WHt

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/056,149	01/24/2002	Nobuyasu Itoh	JP9-2000-0337 (291)	2291

40987 7590 02/04/2005

AKERMAN SENTERFITT
P. O. BOX 3188
WEST PALM BEACH, FL 33402-3188

EXAMINER	
ALBERTALLI, BRIAN LOUIS	

ART UNIT	PAPER NUMBER
2655	

DATE MAILED: 02/04/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	10/056,149	ITOH ET AL.
	Examiner	Art Unit
	Brian L Albertalli	2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-20* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☒ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *7/12/02*.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Specification

1. The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed.

2. The following title is suggested: --Speech Recognition Apparatus and Method Utilizing a Language Model Prepared for Expressions Unique to Spontaneous Speech--.

Claim Objections

3. Claim 3 objected to because of the following informalities: in line 3 of the claim, "included in" should be –including--. Appropriate correction is required.

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. Claims 1-9, 11, 13-15, and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gillick et al. (U.S. Patent 6,167,377), in view of Siu et al. (*Modeling Disfluencies in Conversational Speech*).

In regard to claims 1 and 7, Gillick et al. disclose a speech recognition apparatus and a corresponding method comprising:

a transformation processor (Fig. 2, recognizer 215) configured to transform at least one phoneme sequence included in speech into at least one word sequence, and to provide, for said word sequence, an appearance probability indicating that said phoneme sequence originally represented said word sequence (recognizer 215 generates first language model scores for a set of candidate words using a set of language models wherein the scores are generated by each language model individually, column 16, lines 3-8);

a renewal processor (recognizer 215) configured to renew said appearance probability, provided for said word sequence by said transformation processor, based on a renewed numerical value indicated by language models corresponding to said word sequence provided by said transformation processor (recognizer 215 combines the language model scores for each word to produce a combined score for each word, column 16, lines 8-11); and

a recognition processor (recognizer 215) configured to recognize speech by selecting one of said word sequences for which the renewed appearance probability is the highest to indicate that said phoneme sequence originally represented said selected word sequence (recognizer 215 uses the combined score to determine candidates words that most closely match a user's utterance, column 16, lines 21-25);

wherein said renewal processor calculates said renewed numerical value using a first language model, and a second language model, which differs from said first language model, and employs said renewed numerical value to renew said appearance probability (the combined score is based on a first language model and a different

second language model that produce scores P_1 and P_2, respectively, that are combined to form a new score P_c, which is used as the probability the candidate words were spoken, column 16, lines 3-26).

Gillick et al. further disclose different numbers and types of models can be combined (column 18, lines 30-31).

Gillick et al. do not disclose that the first language model is a language model especially prepared for expressions unique to spontaneous speech.

Siu et al. disclose a language model (page 387, equation 3) that is prepared for expressions unique to spontaneous speech (conversational speech markers, see Table 1, are included in the model, page 388, 2nd column, 2nd paragraph, lines 1-4 and 3rd paragraph, lines 1-2).

It would have been obvious to one of ordinary skill in the art at the time of invention to use the language model as disclosed by Siu et al. as the first language model in the system of Gillick et al., because the language model prepared for expressions unique to spontaneous speech performs about 2.5% better than the baseline trigram language model, as taught by Siu et al. (page 388, 2nd column, 2nd paragraph, lines 4-5 and Table 6).

In regard to claim 2, the language model disclosed by Siu et al., used in the combination of Gillick et al. and Siu et al., as applied to claim 1, above, necessarily determines a probability whether a word sequence includes predetermined words unique to spontaneous speech (the conversational markers given in Table 1), or does

not (which is equivalent to the probability that the word sequence is the originally represented phoneme sequence).

In regard to claim 3, Gillick et al. discloses the renewal processor renews the appearance probabilities of the word sequence based on the first language model and the second language model (the combined score is based on a first language model and a different second language model that produce scores P_1 and P_2, respectively, that are combined to form a new score P_c, which is used as the probability the candidate words were spoken, column 16, lines 3-26). Whichever word sequence is determined as being most probable is converted to a word. Therefore, in the combination of Gillick et al. and Siu et al., as applied to claim 1, above, if a word sequence containing a predetermined spontaneous speech word was determined to be the highest scoring hypothesis, it would necessarily be included in the word sequence.

In regard to claim 4, the language model disclosed by Siu et al., used in the combination of Gillick et al. and Siu et al., as applied to claim 1, above, employs, as an element, a word set including a disfluency (see Table 1, *uh, um, oh*).

In regard to claim 5, Gillick et al. disclose said first and said second language models are defined as N-gram models, and wherein said renewal processor employs, as said renewed numerical value, the weighted average value of said first and said second language models (the language models are unigram, bigram, or trigram models,

column 2, lines 1-5; the first and second language models are weighted by interpolation weights to determine the combination score P_c, column 16, lines 12-19).

In regard to claim 6, Gillick et al. disclose a computer system comprising the speech recognition apparatus (Fig. 1, 125).

In regard to claim 8, the language model disclosed by Siu et al., used in the combination of Gillick et al. and Siu et al., as applied to claim 7, above, the language model is an appearance probability including words unique to spontaneous speech with a combination of N consecutive words (see Table 1, equation 3, and page 388, 2nd column, 2nd paragraph, lines 1-4; an n-gram model generates the appearance probability with a combination of N consecutive words).

In regard to claim 9, in the language model disclosed by Siu et al., used in the combination of Gillick et al. and Siu et al., as applied to claim 7, above, the word unique to spontaneous speech is a disfluency (see Table 1, *uh, um, oh*).

In regard to claims 11 and 17, Gillick et al. disclose a program that permits a computer to perform (column 18, lines 42-43):

acoustically analyzing speech data and transforming said speech data into a feature vector Fig. 3, column 3, lines 33-36);

generating acoustic data, for which an appearance probability is provided, for at least one phoneme sequence that may correspond to said feature vector obtained from said analyzing step (acoustic scores are generated, column 13, lines 20-21);

transforming said phoneme sequence into at least one word sequence (recognizer 215 generates first language model scores for a set of candidate words using a set of language models wherein the scores are generated by each language model individually, column 16, lines 3-8);

renewing said appearance probability by referring to a language model that is written by correlating the appearance probability of at least one word sequence with a combination of N consecutive words (recognizer 215 combines the n-gram language model scores for each word to produce a combined score for each word, column 16, lines 8-11 and column 2, lines 1-5); and

speech recognizing said speech data by using as a speech recognition result one of said word sequences for which the renewed appearance probability is the highest (recognizer 215 uses the combined score to determine candidates words that most closely match a user's utterance, column 16, lines 21-25).

Gillick et al. further disclose different numbers and types of models can be combined (column 18, lines 30-31).

Gillick et al. do not disclose that one of the language models includes disfluencies.

Siu et al. disclose a language model that is an appearance probability including words unique to spontaneous speech with a combination of N consecutive words (see

Table 1, equation 3, and page 388, 2nd column, 2nd paragraph, lines 1-4; an n-gram model generates the appearance probability with a combination of N consecutive words).

It would have been obvious to one of ordinary skill in the art at the time of invention to use the language model as disclosed by Siu et al. to renew the appearance probability in the system of Gillick et al., because the language model prepared for expressions unique to spontaneous speech performs about 2.5% better than the baseline trigram language model, as taught by Siu et al. (page 388, 2nd column, 2nd paragraph, lines 4-5 and Table 6).

Furthermore, Gillick et al., as modified by Siu et al. would necessarily transform a phoneme sequence into a word sequence while a disfluency was included as a word choice selection, because in the transforming step above, the Siu et al. model would produce disfluencies as the highest scoring word choice if they were in the original spoken speech.

Still further, since Gillick et al. disclose that the user can dynamically adjust which words are included in the active vocabulary (column 15, lines 40-47). Therefore, Gillick et al., as modified by Siu et al. would necessarily allow the user to select whether a disfluency was to be reflected in the recognition result.

In regard to claim 13, Gillick et al. disclose outputting said word sequence to which the highest appearance probability applies as text data (column 14, lines 54-57).

In regard to claim 15, Gillick et al., as modified by Siu et al., as applied to claim 10 above, would necessarily renew said appearance probability by referring to said disfluency language model and to a general-purpose language model.

In regard to claim 19, both Gillick et al. and Siu et al. disclose the models are N-gram models (in Gillick et al. the language models are unigram, bigram, or trigram models, see Gillick et al. column 2, lines 1-5; in Siu et al. see Table 1, equation 3, and page 388, 2nd column, 2nd paragraph, lines 1-4). Additionally, Gillick et al. disclose the appearance probability is renewed by using a weighted average model (the first and second language models are weighted by interpolation weights to determine the combination score P_c, column 16, lines 12-19).

6. Claims 12, 14, and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gillick et al., in view of Siu et al., and further in view of Stolcke et al. (*Statistical Language Modeling for Speech Disfluencies*).

Neither Gillick et al. nor Siu et al. disclose adding a symbol to a word indicating that the word is a disfluency choice and removing the word from the word sequence.

Stolcke et al. disclose a method for automatic disfluency tagging and removal (the cleanup model, section 2.2 and page 408, 1st column, 6th paragraph, lines 1-2).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Gillick et al. and Siu et al. to add symbol a word was a disfluency and removing the word from the word sequence before

outputting the text data, so that the output text data could be more easily understood. Written text rarely contains any disfluencies (such as repeated words, or fillers such as *uh*), and is difficult to understand when those disfluencies are included.

7. Claims 10, 16, and 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gillick et al., in view of Siu et al., and further in view of Chen (U.S. Patent 6,067,514).

In regard to claim 10, Gillick et al. disclose different numbers and types of models can be combined (column 18, lines 30-31).

Neither Gillick et al. nor Siu et al. disclose a third language model which is especially prepared for specific symbols.

Chen discloses a language model especially prepared for specific symbols (punctuation) included in word sequences (column 10, line 65 through column 11, line 3).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Gillick et al. and Siu et al. to include a third model especially prepared for specific symbols, in order to automatically insert punctuation into the output text data, which reduces the amount of editing required for the text data and makes the text more easily readable, as taught by Chen (column 1, lines 44-47).

In regard to claims 16 and 20, Gillick et al. disclose different numbers and types of models can be combined (column 18, lines 30-31).

Neither Gillick et al. nor Siu et al. disclose transforming the phoneme sequence into a word sequence while a pause is included as a punctuation choice, and renewing said appearance probability by further referring to a punctuation language model that is limited to punctuation insertion.

Chen discloses a punctuation language model that is limited to punctuation insertion (column 10, line 65 through column 11, line 3).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Gillick et al. and Siu et al. to include a pause as a punctuation choice (for inserting, for example, a period), and renewing said appearance probability by referring to a punctuation language model, in order to automatically insert punctuation into the output text data, which reduces the amount of editing required for the text data and makes the text more easily readable, as taught by Chen (column 1, lines 44-47).

Conclusion

8. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Schultz et al. (*Acoustic and Language Modeling of Human and Nonhuman Noises for Human-to-Human Spontaneous Speech Recognition*), Rose et al. (*Modeling Disfluency and Background Events in ASR for a Natural Language Understanding Task*), and Siu et al. (*Variable N-Grams and Extensions for*

Conversational Speech Language Modeling) disclose various methods for modeling speech disfluencies. Kemp et al. (*Modeling Unknown Words in Spontaneous Speech*) and Suhm et al. (*Towards Better Language Models for Spontaneous Speech*) disclose language models that handle out of vocabulary words associated with spontaneous speech. Lickley et al. (*On no Recognizing Disfluencies in Dialogue*) disclose certain types of disfluencies can aid in speech recognition. Nishimura et al. (U.S. Patent 6,778,958) disclose a method to insert punctuation into text derived from a speech recognizer. Fung (U.S. Patent Application Publication 2003/0023437) discloses a disfluent speech language model. Lee et al. (U.S. Patent Application Publication 2002/0087315) and Nakamura et al. (U.S. Patent 5,875,425) disclose systems that utilize several language models. Imai et al. (U.S. Patent 6,393,398) disclose a method that renews a first language model result with a second, more complex language model. Deligne et al. (U.S. Patent 6,314,399) disclose a method for creating N-gram language models.

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (703) 305-1817. The examiner can normally be reached on Mon - Fri, 8:00 AM - 5:30 PM, every second Fri off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on (703) 305-3011. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 1/31/05

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER